Section 8.2 - Estimating a Population Proportion (pp.492-507)   S(2016)

You guessed it, back to the Hyena Experiment. Suppose on the first day of the semester, your team chose one SRS of size 20 from the bag in order to determine a point estimate $\hat{p}$ for the population proportion p of male hyenas in the Croatan NF. You found 16 male hyenas.

Point estimate $\hat{p} = \frac{16}{20} = 0.8 \Rightarrow \hat{p} = 80\%$

The ultimate goal of today's lesson is being able to construct a confidence interval for a population proportion based upon the point estimate.

---

1. Conditions for Estimating p

Prior to performing confidence interval calculations, 3 conditions must be verified:

(1) Random - ALLOWS INFERENCE ABOUT POP. OR CAUSE/EFFECT.

(2) Normal - ALLOWS FOR COMPUTATION OF A CRITICAL VALUE.

(3) Independent - ALLOWS FOR COMPUTATION OF STAND. DEV.

Example. Using our hyena example, let's check conditions for constructing a confidence interval.

(1) RANDOM: ✓ SRS

(2) NORMAL: $n\hat{p} = 20(.8) = 16 \geq 10$ ✓

$n(1-\hat{p}) = 20(.2) = 4 < 10$ ✗

(3) INDEPENDENT: $n = 20$, IT IS SAFE TO ASSUME POPULATION OF HYENAS $> 10(20) = 200$ ✓

Check Your Understanding - In each of the following settings, check whether the conditions for calculating a confidence interval for the population proportion p are met.

(1) An AP Statistics class at a large high school conducts a survey. They ask the first 100 students who arrive at school one morning whether or not they slept at least 8 hours the night before. 17 students said, "Yes."

RANDOM: NO - CONV. SAMPLE ✗

NORMAL: $n\hat{p} = 100(.17) = 17$ $n(1-\hat{p}) = 100(.83) = 83 \geq 10$ ✓

IND: LARGE HS HAS MORE THAN $10(100) = 1000$ STUDENTS.

(2) A quality control inspector takes a random sample of 25 bags of Hyena brand potato chips from the thousands of bags filled in an hour. Of the bags selected, 3 had too much salt. $3/25 = \frac{12}{100} = 0.12$

RANDOM: SRS ✓

NORMAL: $n\hat{p} = 25(.12) = 3 < 10$ ✗

IND: 1000'S PRODUCED PER HOUR, ∴ 3 < 10% OF POP ✓

2. Constructing a Confidence Interval for *p*

$$\boxed{\textbf{Statistic} \pm \textbf{(critical value)} \bullet \textbf{(standard deviation of the statistic)}}$$

a. Sample Statistic - The *sample statistic* is $\hat{p}$

** In the hyena example, this is $\hat{p} = 0.8$

b. Standard Deviation - The standard deviation of the sampling distribution of $\hat{p}$ is $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$

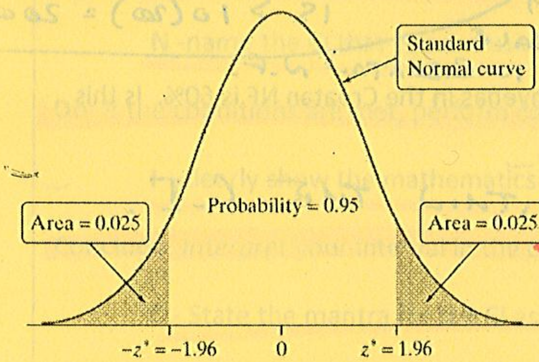
Since we do not know the value of p, we replace it with the sample proportion $\hat{p}$ $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

This quantity is known as the ***standard error (SE)*** of the sample proportion $\hat{p}$. When the standard deviation of a statistic is estimated from the data, the result is called the **standard error** of the statistic.

Note: The formula for the standard error of the sample proportion is not on the AP Exam Formula sheet but the formula for the standard deviation of the sampling distribution of $\hat{p}$ is. Merely substitute $\hat{p}$ into the formula to compute the standard error of the sample proportion.

** In the hyena example, this is $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.8(0.2)}{20}} \approx \boxed{0.089}$

c. Critical Value - How do we get the critical value for our confidence interval? For our "mystery mean" in Section 8.1, we used a critical value of 2 based on the 68-95-99.7 rule for Normal distributions. We can get more accurate results with Table A or a calculator.

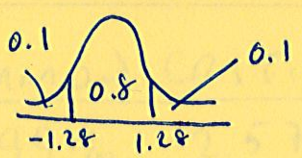


In order to find a level C confidence interval, we need to catch the central area C under the *standard Normal curve*. We will use the notation z^* to denote critical z-values.

It is very important that you use this notation and not just z to show critical values.

** For the hyena example, let's find the critical value z^* for an 80% confidence interval.

(1) Table A:



.08

−1.2 $\boxed{.1003}$

(2) Calculator:

INVNORM(0.1) = −1.281551567 ≈ −1.28

$\boxed{z^* = 1.28}$

WE NEED ± 1.28 STD DEV'S TO HAVE 80%.

With the three components of the confidence interval addressed, we can now put it together in what is called a **One-Sample z Interval for a Population Proportion.**

One-Sample z Interval for a Population Proportion

Choose an SRS of size n from a large population that contains an unknown proportion p of successes. An approximate level C **confidence interval for p** is

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

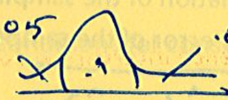
NOT ON FORMULA SHEET (DISCUSS)

where z^* is the critical value for the standard Normal curve with area C between $-z^*$ and z^*. Use this interval only when the numbers of successes and failures in the sample are both at least 10 and the population is at least 10 times as large as the sample.

Example. Suppose another team chose an SRS of size 20 from their bag of hyenas and got 10 males.

(1) Calculate and interpret a 90% confidence interval for p.

STATISTIC: $\hat{p} = 0.5$



$z^* = 1.64$

$SE = \sqrt{\frac{0.5(0.5)}{20}} = 0.112$

$0.5 \pm (1.64)(0.112) =$

$[0.316, 0.684]$

RANDOM: SRS ✓
NORMAL: $n\hat{p} = 20(.5) = 10$
$n(1-\hat{p}) = 20(.5) = 10$ ✓
IND: ASSUME POP OF HYENAS IS $> 10(20) = 200$ ✓

WE ARE 90% CONFIDENT THAT INTERVAL FROM 0.316 TO 0.684 CAPTURES THE TRUE PROP. OF MALE HYENAS IN CROATAN N.F.

(2) Suppose a claim was made that the proportion of male hyenas in the Croatan NF is 60%. Is this proportion plausible?

YES, SINCE 0.6 IS WITHIN THE C.I.

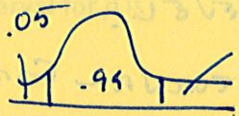
Check Your Understanding - Complete CYU on p. ~~490~~ 499.

① PARAMETER: p = TRUE PROPORTION OF ALL U.S. COLLEGE STUDENTS WHO ARE CLASSIFIED AS ~~BED~~ FREQ. BINGE DRINKERS.

② RANDOM: RANDOMLY CHOSEN ✓

IND: REASONABLE TO ASSUME MORE THAN 10 (10,904) STUDENTS

NORMAL: $n\hat{p} = 10{,}904\left(\frac{2486}{10{,}904}\right) \geq 10$ AND $n(1-\hat{p}) = 10{,}904\left(\frac{8418}{10{,}904}\right) \geq 10$ ✓

③ 

INVNORM (.05) = −2.576 SO $z^* = 2.576$

CI: $\left[0.228 \pm 2.576\sqrt{\frac{0.228(1-0.228)}{10904}}\right] = (0.218, 0.238)$

④ WE ARE 99% CONF. THAT THE INTERVAL (0.218, 0.238) CAPTURES THE TRUE PROPORTION OF U.S. COLLEGE STUDENT WHO ARE CLASSIFIED AS FREQUENT BINGE DRINKERS

3. Putting It All Together: The (Vaunted) Four-Step Process

Confidence Intervals: A Four-Step Process

State: What *parameter* do you want to estimate and at what *confidence level*?

P - parameter of interest

→ 1-SAMPLE Z-INTERVAL

Plan: Identify the appropriate *inference method*. Check *assumptions* and *conditions*.

A - assumptions/conditions - Random, Normal, Independent?

N -name the CI that you are using

Do: If the conditions are met, perform *calculations*.

I - clearly show the mathematics behind creating the interval

Conclude: *Interpret* your interval in the context of the problem.

C - State the mantra for the CI as your conclusion

(Remember - when faced with a confidence interval problem, do not **PANIC!**)

COMMON CRITICAL VALUES:

99%	2.576
95%	1.96
90%	1.645
80%	1.28

Example - According to an article in the *San Gabriel Valley Tribune* (Feb 13, 2003), "Most people are kissing the 'right way.'" That is, according to the study, the majority of couples tilt their head to the right when kissing. In the study, a researcher observed a random sample of 124 couples kissing in various public places and found that 83/124 (66.9%) of the couples tilted to the right. Construct and interpret a 95% confidence interval for the proportion of all couples who tilt their heads to the right when kissing.

"PANIC"

① STATE: WE WANT TO ESTIMATE p = THE TRUE PROPORTION OF COUPLES WHO TILT THEIR HEADS TO THE RIGHT WHEN KISSING AT THE 95% LEVEL.

P ✓

② PLAN: WE WILL USE A 1-SAMPLE Z INTERVAL FOR p

N ✓

CONDITIONS: RANDOM - RANDOM SAMPLE ✓

NORMAL - $n\hat{p} = 124(.669) \approx 83 \geq 10$

$n(1-\hat{p}) = 124(.331) \approx 41 \geq 10$ ✓

INDEPENDENT - # OF COUPLES IN POP. IS AT LEAST 10(124) = 1240 ✓

A ✓

③ DO: $\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.669 \pm 1.96 \sqrt{\frac{0.669(1-0.669)}{124}}$

$(z^* = 1.96)$ $= 0.669 \pm 0.083 = [0.586, 0.752]$

I ✓

④ CONCLUSION: WE ARE 95% CONFIDENT THAT THE INTERVAL FROM 0.586 TO 0.752 CAPTURES THE TRUE PROPORTION OF COUPLES WHO TILT THEIR HEADS TO THE RIGHT WHEN KISSING.

C ✓

Technology - The calculator can be used to construct a confidence interval for an unknown population proportion. Refer to p. 64 of NTA or p. 492 of the text.

[STAT] <TESTS> A: 1PROPZINT X: 83

n: 124

C-LEVEL: .95

(0.58655, 0.75216)

It should be noted that if you use the calculator, it is recommended that you check your answer with the calculations of the formula.

4. Choosing Sample Size - In planning a study, we may want to choose a sample size that allows us to estimate a population proportion with a given margin of error. Essentially the way this works is to make a guess as to what $\hat{p}$ is and then work "backwards." Often times, $\hat{p}$ = 0.5 is used as a guess because it will yield the largest ME.

Sample Size Determination for Desired Margin of Error

To determine the sample size n that will yield a level C confidence interval for a population proportion p with a maximum margin of error ME, solve the following inequality for n: C **confidence interval for *p* is**

$$z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq ME$$

where $\hat{p}$ is a guessed value for the sample proportion. The margin of error will always be less than or equal to ME if you take a guess of 0.5

TYPO TYPO

Example - Suppose that you wanted to estimate p = the true proportion of students at WCHS who have a tattoo with 95% confidence and a margin of error of no more than 10%. How many students should be randomly surveyed to estimate p within 0.10 with 95% confidence?

NO PREVIOUS KNOWLEDGE $\Rightarrow \hat{p} = 0.5$ $\quad ME = 0.10$

$z^* = 1.96$

$1.96\sqrt{\frac{0.5(1-0.5)}{n}} \leq 0.1$

$\sqrt{\frac{0.5(0.5)}{n}} = \frac{0.1}{1.96}$

$\left(\frac{0.1}{1.96}\right)^2 = \sqrt{\frac{0.5(0.5)}{n}}^2 =$

$\left(\frac{0.1}{1.96}\right)^2 = \frac{.25}{n}$

$n = \frac{.25}{\left(\frac{0.1}{1.96}\right)^2} = 96.04$

WE NEED 97 STUDENTS.

Check Your Understanding - Complete CYU on p. ~~484~~ 503

1) $1.96\sqrt{\frac{0.8(0.2)}{n}} \leq 0.03 \Rightarrow n = 683$

2) $n \uparrow$ $\quad 2.576\sqrt{\frac{0.8(0.2)}{n}} \leq 0.03 \Rightarrow n = 1180$

HW: Read Sec 8.2; do problems 27, 29, 35, 36, 39, 44, 45, 49-52